

Masterarbeit

Konversationsbasierte Suche mittels RAG im Simulationsdaten-Management

Bei der Entwicklung moderner Produkte stützen sich Ingenieurteams auf komplexe physikalische Modelle und Simulationen, um bereits virtuell am Computer Konzepte schnell zu testen, optimieren und abzusichern. Dabei entstehen insbesondere für komplexe Produkte große, heterogene Datenmengen, die in einem Simulationsdatenmanagement (SDM) gebündelt, strukturiert und für verteilte Teams kollaborativ nutzbar gemacht werden.

Um die Daten über ein Chat-Interface konversationsbasiert zugänglich zu machen, wurde in SCALE.sdm bereits eine grundlegende Retrieval-Augmented-Generation-(RAG-) Pipeline aufgesetzt, die natürliche Spracheingaben mit im SDM-System gespeicherten Wissensbeständen verbindet. Aufbauend auf dieser Basis soll die Arbeit die SDM-relevanten Anwendungsfälle systematisch identifizieren und untersuchen sowie die Integration von Knowledge-Graphen in das bestehende RAG-System konzipieren, um kontextuelle und präzisere Antworten zu erzielen. Hierfür ist die RAG-Pipeline zielgerichtet anzupassen und zu erweitern. Dazu gehört auch ein dafür geeignetes Indexierungs- und Vektorablagekonzept (Vector Store, Metadatenmodell, Chunking/Embedding-Strategie) sowie deren Validierung im Betrieb.

Der wissenschaftliche Schwerpunkt liegt auf der Graph-gestützten Erweiterung von RAG: Entitäten, Relationen und Domänenwissen sollen über Knowledge-Graphen nutzbar gemacht und in Retrieval, Kontextaufbereitung und Antwortgenerierung integriert werden.

Mögliche Aufgabenstellungen im Rahmen dieser Masterarbeit könnten sein:

- Einarbeitung in SCALE.sdm, SDM-Prozesse und die vorhandene RAG-Pipeline
- Use-Case-Discovery im SDM: Erhebung, Priorisierung und Konkretisierung relevanter Anwendungsfälle (z. B. Suche in Simulationsinputdaten und Änderungsdokumentation, Modell- und Prozessdokumentation und Simulationsergebnissen/Reports sowie allen dazugehörigen Metadaten)
- Datenaufbereitung und Indexierung: Definition eines Dokument- und Metadatenmodells, Extraktion/Normalisierung, Chunking, Embedding-Strategie; Anlegen/Anpassen der notwendigen Indizes und Vector Stores
- Knowledge-Graph-Integration:
- Modellierung relevanter Entitäten/Relationen (z. B. Projekt – Modell – Run – Ergebnis – Testfall – Anforderung)
- GraphRAG-Strategien evaluieren (Subgraph-Retrieval, Pfadbegründungen, Knowledge-Graph-gestützte Filter/Booster im Retriever, Query-Routing)

- Vergleich RAG vs. Knowledge-Graph-augmentiertes RAG und Ablationsstudien.
- Test und Bewertung: Definition von Metriken, Nutzerstudien mit SDM-Anwendern, Skalierungs- und Robustheitstests

Was du mitbringen solltest

- Du befindest dich im Masterstudium (z. B. Informatik, Data Science, Computational Engineering o. Ä.).
- Solide Python-Kenntnisse und grundlegende Erfahrung mit Datenverarbeitung und Git.
- Interesse an KI/LLMs und Suchsystemen sowie Lust auf praktische Umsetzung in einer RAG-Pipeline.

Was auch gut wäre

- Erste Erfahrungen mit LLM-Tools (z. B. OpenAI/Hugging Face) oder Frameworks (LangChain/LlamaIndex).
- Basiswissen zu Knowledge Graphs - oder die Bereitschaft, dich in dieses Thema einzuarbeiten.

Diese Masterarbeit ist ideal für Studierende mit Interesse an KI und Ingenieurwesen, die praxisnahe Innovationen in einem industriellen Umfeld vorantreiben wollen und bietet die Möglichkeit, eine bestehende RAG-Suche im industriellen SDM-Umfeld gezielt auszubauen, praxisrelevante Use Cases zu identifizieren und umzusetzen und durch Knowledge-Graph-Integration wissenschaftlich fundierte Verbesserungen bei Kontextualität, Nachvollziehbarkeit und Antwortqualität zu erreichen. Die Masterarbeit wird in Zusammenarbeit mit der Abteilung für Berechnungsmethoden bei Volkswagen (VW) sowie dem Fraunhofer-Institut SCAI (Numerical Data-Driven Prediction, Institute for Numerical Simulation) bzw. einem anderen Lehrstuhl mit Bezug zu deiner Hochschule betreut.

Idealerweise arbeitest du vor Ort in unserem Büro in Dresden und erhältst eine direkte 1:1-Betreuung durch unsere Softwareentwickler sowie Zugriff auf die nötigen Hardwareressourcen. Alternativ ist die Durchführung an einem unserer anderen SCALE-Standorte möglich. Die Masterarbeit wird im üblichen Rahmen vergütet.

Interessiert? Bewirb dich per E-Mail an jobs@scale.eu mit Lebenslauf, wenn vorhanden Link zu einem eigenen GitHub-Repo und Notenspiegel.